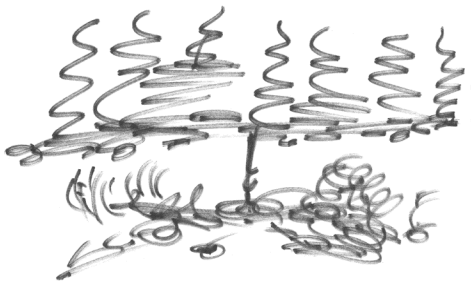

The Mountain-Whisper-Light Statistics
Nayak L. Polissar, PhD
1827 23rd Ave. East, Seattle, WA 98112-2913
Phone: (206) 329-9325 • Fax: (206) 324-5915
E-mail: nayak@mwlight.com

Date: 4/1/13
To: Michael Opheim
From: Nayak Polissar
Re: "Seldovia Report"

Dear Michael,

It was my pleasure to read the report by you and your colleagues. It is amazing that you have accomplished so much with limited resources. Plus, in the negotiations about environmental regulation, all of the stakeholders will appreciate having some data to rely on.

I apologize that I could not do a thorough, line-by-line review. I have done what I have time for. So, the points that I have made are not the only points to be considered as a review of this report. Lon Kissinger has offered extensive comments, for example.

I feel that it important that the report be revised in order to get the attention and respect that the survey deserves.

Here are some general comments.

The quality of writing is very good in many places.

The report feels much too long to me, and a lot of that feeling comes from reading text of lists of numbers that should be in tables. Some things could be moved to appendices, too.

The description of methodology shows some confusion between sampling people and sampling households.

The description of statistical methodology is hard to follow.

Many of the pie charts do not follow standard pie chart methodology, and probably should not be pie charts. Use bar charts or tables or something else.

The minimal attention given to shellfish is puzzling.

There is a general lack of clarity in many places. More attention (but not more space) needs to be given to defining and explaining.

I do recommend a thorough revision and that you work with an experienced statistician (or at least an experienced quantitative person) who has worked on large quantitative reports. Each major section should be examined to determine what message needs to be communicated; the section can be revised with that in mind.

Best wishes,

Nayak

Specific comments

Top of page 7:
Add more references for surveys in Oregon and Washington.

Section 5.2.2:
This is a nice section. Very compelling.

6.1.1 & 6.1.2:
You describe the sampling frame, but what is your target population. About whom (what population) is the survey? What residential area or villages, what Tribes or ethnic groups, what are the "rules" for membership in that population, etc.

Describe in detail the random selection process. (You may have covered random selection later.) You mention selecting at most one per household, which would underrepresent individuals living in larger HH. How was selection done? Was a HH "closed out" when you had one adult member from the HH? You can't undo this now, but it should be discussed in your discussion section.

"Selection of tribal members participating in the assessment was random thus eliminating any potential bias from interviewers." Interviewers can introduce bias during the interview. It seems like you may be saying that the interviewers or some person did not do the selection. Clarify. You may have avoided selection bias but any survey done by interviewers will have some degree of interviewer bias, hopefully small.

"However, whenever the data were compiled, the data were weighted based upon the number of tribal households in each village." You selected the sample based on people (it seems) but weighted (to get a population total figure) based on HH. Apples and oranges. It will be OK if the number of persons per HH is the same in each village, or

approximately so. If you can't fix it, these kinds of methodologic errors don't invalidate the survey, but they make it harder to interpret. Everything starts to become approximate. If you can fix it, then fixing it would be good. If you can't then it would be good to make (non-embarrassing) statements about it.

On page 8 or somewhere, it would be good to show the numerical weights used to combine villages to yield a population figure.

Frankly, I am not very concerned about the basis for choosing the sample size. It is nice to know it, but that section could be shortened considerably. Whether the planning was fantastic or terrible, it doesn't matter, as long as the sampling and interviewing, etc., was done in an unbiased way. Also, rather than breaking the sample size equation into parts (pages 9 and 10), it would be better and more in line with convention (and more referenceable) to have just one equation that takes account of the finite population, the SD and other parameters. I did not check your equations. You should be able to derive the required sample size with one equation, definition of terms in the equation and a few words.

Also, I believe that you wanted to get precise estimates of rates of consumption and not of proportions. Some of your sample size equations and sample sizes are based on getting a certain precision of proportions (or percentages), but then you bring in consumption rates (I think) in the equation in the middle of page 10.

Frankly, the whole sample size section is a bit of a head-scratcher. After investing some frustrating time on it, I have run away from it. It needs a total re-do.

The standard deviation estimate section is also too long, and it is not clear how you used Table B-3; it can probably be dropped. At this point (now) having chosen 30 grams does not matter, but it does appear that you used limited resources to make that choice. How about just saying that after consideration of various sources, you chose 30g as the SD?

6.1.3:
From page 10-11: "Tribal members who could not be contacted after a minimum of four attempts, or refused to participate, were removed from the sample set and were replaced by the next eligible members down the list following the same selection method as above." How was the list sorted or organized? If it was sorted by HH, then people in an HH with a person not responding might be more likely to be selected than people in a HH with responsive people. The best thing would be if it was sorted randomly. Details needed on how the list was sorted and what taking the next member down the list really means.

I am putting this comment here, but somewhere you need to give a through report on response rates, preferably per village. How many people had contact attempts in each village? How many were actually contacted and invited to participate? How many refused or could not participate? How many were interviewed? The very minimum is to

tell the number of people selected from your list for contact (initially or as a replacement) and then the number of those who were interviewed. Express that as a percentage response rate, too.

6.1.4:
In this section it appears that you are weighting by HH, even though your selection was of individuals. See my comment above.

We can discuss it by phone, but I believe that your standard errors are calculated without reference to the clustering by villages. And, if the four villages are the "population" that you are wanting to provide rates for, then your SEs are too big.

This is a complex topic that we can discuss in the teleconference, but a lot depends on what population we are talking about. The tow choices I see are:

a) this survey is intended to provide rates for just these four villages considered as a population,

or,

b) the villages may be considered as some sort of random or representative sample of villages which are part of a larger population of villages, perhaps the whole Tribe or the portion resident with access to this water resource.

6.2.1:
The choice of the youngest child member of the HH means that the sample of children is not representative of children in the villages. The population sampled is the youngest children in HH. That may be the most vulnerable group, and the rates for that group are valuable in that way, but we can't use the summary rates for that group to apply to children in general. They will be useful rates. but in an advisory or illustrative way. Somewhere you should show the age and gender distribution of the children sampled.

Consumers/non-consumers. Did you keep track of this per species or per species group?

6.2.2:
I suggest including the questionnaire as an appendix to this document. The questionnaire is extremely important in understanding the data.

6.3.1:
24-hour recall. I believe that you probably asked the people what they ate during the entire preceding day, right? I don't think that you asked them about the consumption during the 24-hours preceding the specific clock time that the question was asked during the interview. Re-write. This kind of wording also happens I section 7.2.3.

6.3.2.1:
I have no idea what you mean by "seasonal…correlations in consumption".

6.3.2.3:

It is too bad that shellfish were not given the same attention (for quantitative rates) as finfish. No models for shellfish consumption? Shellfish live in the water, too.

6.5:
76 interviews were completed. Again, what was the response rate? 76 out of how many people with attempted contacts?

6.6.2:
How many pretests? How many people had pretests? Ideally, those pretest results should not be used in the final analysis and the people who did the pre-tests should not be included in the sample that was analyzed.

I note that you included non-consumers in your consumption rates. There is interest in having rates for consumers-only as a population, so I do suggest that for major species groups (all species, finfish, etc.) you have both per capita rates (including non-consumers) and consumer-only rates.

To me it doesn't make sense to compare your rates and findings only or mainly with the Columbia River Tribal populations. As Lon pointed out in his review, there are other Tribes in WA/OR that have relevant rates for comparison. Yes, you may have used a methodology similar to the CRITFC surveys, but if I measure the height of a chair and of a doorway with a tape measure (i.e., using the same measuring instrument), it doesn't mean that the doorway is the best thing to compare to a chair.

6.8.4:
Outliers: "Due to the small sample sizes, no values were excluded from analysis." That doesn't make sense. There may be good reasons not to exclude values, but small sample size is not a good reason. Here is a set of numbers:
2
6
5
3
7
4
9
700

The sample size is small, but clearly something must be done about the last value before calculating a representative number, such as a mean.

7.1.2:
Table 2 shows weighted percentages, but, as stated earlier, it would be nice to know what numerical weights you are using per village. Figure three has relatively narrow age categories, given the sample size. Who about using the age categories of Table 2 for the Figure. Also, it would be nice to have all four villages in Table 2 along with the combined count. You have space in the table for that. The mean (and SE) values in the

paragraph above Table 2 could also be put in a table. This is an example, repeated many times in the report, of having to read what should be a table as a list in text.

7.2:
Suggest you put values from Figure 4 in a table instead of a Figure. Again, why compare to just CRITFC?

Re: "The average rate of consumption by all interviewed adults (n=76) throughout the year for all species from all sources was determined to be 94.8 (±23.55 SE) grams per day (g/d) (Figure 4)." With this SE, it looks like the SD would be at least 184[1] g/day (larger than the mean), which suggests that the data are highly skewed to the right and/or have some exceptionally large values. It would be nice to see a graphical display of the individual consumption rates, such as a cumulative distribution or a histogram.

7.2.1.3:
The SEs in the first two lines of this section imply very large SDs

Re: "Surprisingly, the standard deviation calculated from the g/d data was higher than what we originally anticipated when calculating the sample size needed for the assessment." A probable reason for this is that your SD calculations include the village-to-village variation along with the person-to-person variation within a village. When we decide whether the four villages (alone) are the population of interest or not, then we can calculate appropriate SDs and SEs.

Notation: "$p \geq 0.05$". Here and in many places you use this kind of notation. It is best to quote the actual p-value. It is useful. For example, here, $p = 0.07$ would still be considered "marginally significant", given this sample size, whereas $p = 0.8$ would not be. In general it is good to quote specific p-values, but it is fine to have a bound for very small p-values, such as $p < 0.001$.

This section is also another example of the need to put a whole series of consumption rates in a table rather than list them in text. This happens in a number of sections.

Figure 5 would be better in a Table. I (and many others) have a black and white printer, and the colors don't contrast well in grey shades. Also, given the small sample size per village, some of the rates are based on a handful of people. The differences that you see in the bars are heavily influenced by noise (random variation.)

Outlier? In Figure the huge SE bar for Port Graham's 40-59 age group suggests that there is one person—maybe 2—who have a really, really high consumption rate. Please check it out.

7.2.2:

[1] I am multiplying the SE by approximately the square root of the sample size, 76, to get an approximate SD.

Re: “For all months identified as high fish consumption months by the entire population sampled….” This is confusing. Please clarify how you chose the specific months and which months they are. Is it June-August? The top paragraph on the next page is also confusing. How were the specific “low” months selected? Are they November-May? What does the 16/76 = 20.8% refer to (top of page 26)? Is it the percentage citing January as a low-consumption month?

7.2.3:
It would be helpful to see the Spearman correlation coefficient (and p-value) between the 24-hour recall and the overall consumption rates.

7.2.4:
Here and many other places, you should give rates to at most one decimal place. Your precision does not support two decimal places.

7.2.5:
A table would be helpful here and could replace some of the text.

Figures 9-12. I suggest you drop the figures and show a table of rates and percentages. Order the rows of the table from high to low consumption rates.

7.2.6:
The percentages are not really helpful or meaningful. Let’s discuss this. One way to use the data is to take the most frequently consumed species and show (in a table) the percentage of consumers (of each species) who consume each specified part.

7.3:
I was surprised that “smoked” was the most frequent form of fish consumed. Could this be driven by snacks, which may be frequent but which may not contribute the largest quantity to fish consumption.

I will give my comments on the balance of the report during the teleconference.

Best wishes,

Nayak